NAVAL POSTGRADUATE SCHOOL

Monterey, California

THESIS

SPEECH RECOGNITION APPLICATION IN C.I.C.

by

LCDR Constantinos P. Leventis Hellenic Navy.

September,1991

Thesis Advisor: Gary K. Poock

Approved for public release; distribution is unlimited

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Naval Postgraduate School 6b. OFFICE SYMBOL (*If applicable*) 32	7a. NAME OF MONITORING ORGANIZATION Naval Postgraduate School
6c. ADDRESS (*City, State, and ZIP Code*) Monterey, CA 93943-5000	7b. ADDRESS (*City, State, and ZIP Code*) Monterey, CA 93943-5000
8a. NAME OF FUNDING/SPONSORING ORGANIZATION 8b. OFFICE SYMBOL (*If applicable*)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (*City, State, and ZIP Code*)	10. SOURCE OF FUNDING NUMBERS Program Element No. / Project No / Task No. / Work Unit Accession Number

11. TITLE (*Include Security Classification*)
SPEECH RECOGNITION APPLICATION IN C.I.C.

12. PERSONAL AUTHOR(S) LEVENTIS, CONSTANTINOS P.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (*year, month, day*)	15. PAGE COUNT
Master's Thesis	From To	1991, SEPTEMBER	58

16. SUPPLEMENTARY NOTATION
The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

17. COSATI CODES			18. SUBJECT TERMS (*continue on reverse if necessary and identify by block number*)
FIELD	GROUP	SUBGROUP	CONTINUOUS VOICE RECOGNITION SYSTEM, TACTICAL TABLE, VERBEX SERIES 5000 VERSION 3.00, MANUAL MODE, COMBAT INFORMATION CENTER, ORAL MODE

19. ABSTRACT (*continue on reverse if necessary and identify by block number*)

THE USE OF A CONTINUOUS VOICE RECOGNITION SYSTEM FOR DATA INPUT TO TACTICAL TABLES IN THE COMBAT INFORMATION CENTER WOULD IMPROVE THE MAN-MACHINE INTERFACE AND DECREASE THE REACTION TIME OF OPERATORS WHO RUN THE TABLES. THE RESULTS OF THIS STUDY SHOW THAT THE DELAY TIMES OF TRAINED PERSONNEL USING MANUAL TYPING INPUT METHODS WERE FAR GREATER THAN WHEN THEY USED CONTINUOUS SPEECH INPUT TO RUN TWO TACTICAL TABLES. USING A VERBEX SERIES 5000 VERSION 3.00 CONTINUOUS SPEECH RECOGNITION SYSTEM, THE OPERATORS' REACTION TIMES WERE IMPROVED BY A FACTOR OF 3.3 AND AT THE SAME TIME THEY COMMITTED FEWER DATA ENTRY ERRORS WHEN RUNNING THE TABLES WITH SPEECH INPUT. THE SUBJECTS WHO PARTICIPATED IN THE EXPERIMENTS ALSO SUBJECTIVELY REPORTED THAT THE FREEDOM ALLOWED BY SPEECH INPUT WAS AN IMPROVEMENT OVER MANUAL TYPING INPUT METHODS. USING SPEECH INPUT, ONE OPERATOR COULD RUN TWO TACTICAL TABLES, WHERE NOW IT TAKES TWO TO THREE PEOPLE TO DO THE SAME JOB.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION	
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS REPORT ☐ DTIC USERS	UNCLASSIFIED	
22a. NAME OF RESPONSIBLE INDIVIDUAL POOCK, GARY K.	22b. TELEPHONE (*Include Area code*) (408) 646-2636	22c. OFFICE SYMBOL OR/Pk

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted
All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

Approved for public release; distribution is unlimited.

Speech Recognition Application in C.I.C.

by

Constantinos P. Leventis
Lieutenant Commander, Hellenic Navy
B.S. in Telecommunications Systems Management,
Naval Postgraduate School, 1991

Submitted in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE IN TELECOMMUNICATIONS SYSTEMS MANAGEMENT

ABSTRACT

The use of a Continuous Voice Recognition System for data input to Tactical Table in the Combat Information Center would improve the man-machine interface and decrease the reaction time of operators who run the tables. The results of this study show that the delay times of trained personnel using manual typing input methods were far greater than when they used continuous speech input to run two tactical tables. Using a Verbex Series 5000 Version 3.00 continuous speech recognition system, the operators' reaction times were improved by a factor of 3.3 and at the same time they committed fewer data entry errors when running the tables with speech input. The subjects who participated in the experiments also subjectively reported that the freedom allowed by speech input was an improvement over manual typing input methods. Using speech input, one operator could run two tactical tables, where now it takes two to three people to do the same job.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. EXECUTIVE SUMMARY

This study describes an experiment in which military officers used a Continuous Voice Recognition System, VERBEX Series 5000 Version 3.00, as a data input method for Tactical Tables.

The objective was to prove that when using normal typing input, delay times of even the most trained personnel are greater than those of a person using a Continuous Voice Recognition System as an input method for Tactical Tables.

Ten military officers were introduced to the voice equipment for the first time, trained on the system by creating their own voice patterns and then run the scenario which was given to them.

In the experiment, subjects followed a fixed scenario of instructions which they performed one time with voice input, and then a second time with manual input.

The scenario was designed to take about six minutes to perform with normal manual input.

Most of the subjects (eight out of ten) had some familiarity with other Voice Recognition Systems. All subjects were introduced to the VERBEX Series 5000 Version 3.00 System and instructed on what they would be doing.

The results show:

1) Voice Input was 3.3 times faster than the manual typing

input.

2) Manual Typing Input had 5.69 times more entry errors.

3) Voice Input did not give any erroneous reply, and in case of doubt did not reply at all, in this way requiring the user to repeat his utterance.

We have observed here that with minimal practice and minimal experience with the system, the job was done faster, and with fewer errors.

II. INTRODUCTION

It is well known that today Naval Operations are done with the use of specially trained personnel, due to the complexity of the existing systems such as Tactical Tables to support advanced weapon systems.

The situation in the Persian Gulf, caused by the invasion of Kuwait by Iraq, proved the importance and the need to minimize operator reaction time to obtain the tactical advantage over the enemy. As an example to this we can mention the SKUD missiles launched by Iraq against Israel, and the inability of the last ones to intercept them even though they possessed very sophisticated weapons like the "Patriot" anti-missiles. They could have been more effective if they had been able to reduce their reaction time with the use of a Continuous Voice Recognition System.

Even the most trained personnel cannot eliminate the delay times caused by the "middle man", (a possible source of misinterpretation). This guides us to the use of "real time" systems offered by the new technology, such as Continuous Voice Recognition Systems.

It has been acknowledged that speech is a human being's most effective and therefore most comfortable means of communicating (Strategic Computing Program, Integration,

Transition and Performance Evaluation of Speech Technology, 1985).[Ref. 1]

Computers which can operate via voice commands may be logical alternatives as input devices eliminating most possible sources of misinterpretations, the "middle man".

As LeFever states,"Increased use of computers in problem-solving will demand more emphasis on man-machine interfaces. Speech Recognition will be that interface which makes the computer a true extension of man (LeFever, 1987)." [Ref. 2]

A. SPEECH RECOGNITION TECHNOLOGY

1. Background

The original development of speech Input/Output (I/O) technology can probably be traced to the early 1950's and 1960's when many of the larger technical companies were doing basic research on trying to recognize spoken digits.

A few of the companies involved in those days were IBM, Bell Telephone Laboratories, RCA, Philco-Ford and others. As events unfolded, the first commercially available products came on the market in the early 70,s with the Speech Recognition Systems offered by Threshold Technology, Inc. and Scope Electronics, Inc.

Between 1971-1976, the Advanced Research Projects Agency (ARPA) funded a large $15 million research effort. The goal of this effort was to produce systems which could

interpret or "understand" vocabularies of 1000 words when used in continuously spoken sentences or phrases.

A variety of industrial companies, academic institutions, and institutes worked on what was known as the ARPA Speech Understanding Research (SUR) project. The 1978-1980 period saw a good variety of Speech Recognizers become commercially available for a few hundred dollars up to $20,000 and more. Most products in that era were isolated word (utterance) type systems versus connected speech systems which have become more prevalent in the 1980's. An utterance is one or more words in a phrase (Poock,1984).[Ref. 3]

2. Types of Speech Systems

There are four generic types of Speech Recognition Systems. One delineation is that a recognition system is either speaker dependent or speaker independent.

If a system is speaker dependent, then it would require samples of the potential user's voice to be in memory in order to work properly. A speaker dependent system is basically tuned for the user's voice and should work better than a speaker independent because the first one contains samples of the actual user's voice.

A speaker independent system contains algorithms which supposedly can handle many different voices and dialects. The system should be able to recognize the voice of anyone who tries to use it. Thus it requires no previous samples of a

given user's voice but rather, contains an algorithm which is robust enough to recognize anyone who talks to it. Both systems today achieve accuracies in the 97-99% range for vocabularies of several hundred words (utterances) in a normal office type environment.

The other generic breakdown of Speech Recognizers is whether they are of the discrete isolated word type or if they are a connected (continuous) Speech System. Either one could be dependent or independent.

In a discrete system, the user has a given number of sound patterns in memory. A sound pattern can be one or several words in a continuous phrase of sounds. When using the discrete system, the user must pause about .10 seconds between each utterance he/she makes.

When the Recognizer "hears" the pause, it knows that was the end of an utterance and therefore starts to search in the memory for what was just said.

Connected Speech, on the other hand, requires no pauses between utterances or phrases.

A slight distinction is made by some in the speech community between Continuous Speech versus Connected Speech.

Connected Speech, allows the user to speak in a natural manner while the computer stores the spoken words in a buffer memory. When the speaker pauses for a breath or between phrases, the information in the buffer is presented to the processors for recognition and appropriate action.

Continuous Speech also allows the speaker to talk in a natural manner; the fundamental difference is that the system continually recognizes what is being said and responds accordingly. Continuous Systems are most natural and comfortable for the user in an interactive environment, but they require more advanced technology (Poock,1984).[Ref. 3]

3. Voice Recognition, Verification, and Identification

Voice Recognition means that we are simply trying to recognize the pattern of sound that was uttered.

Voice Verification means that the user identifies himself by some mean like a Personal Identification Number (PIN) or some similar technique, and than the system verifies "it is/it is not" the real user.

In Voice Identification the entire data base is searched to try to identify the speaker. This is much harder, because the user simply asks to be identified. The process may take longer, but the advantage is that the user need not remember or enter any password or PIN (Poock,1984).[Ref. 3]

B. VERBEX SERIES 5000 VERSION 3.00 VOICE RECOGNIZER

The VERBEX Voice I/O System is a computer peripheral that allows users to send data to computers by voice. In many computer applications, the Recognizer will work as a replacement for a keyboard, leaving the operator's hands and eyes free (VERBEX Grammar Development Manual, 1990).[Ref. 4]

In order to understand what the VERBEX Recognizer does, think of the Recognizer as a translator which performs much the same function for a Host computer as a human translator does for a foreigner. The Recognizer translates words spoken through the headset microphone into computer code the Host computer can understand, and then sends this information to the computer. This process is called Recognition.

Understanding and translating spoken language into digital information is a truly monumental feat for any machine. To accomplish this task, the Recognizer contains sophisticated computers of its own. Before the Recognizer can begin recognition, it's internal processors must be given two banks (files) of information:

1. **Recognizer File**

This Recognizer File contains the following:

a. A list of the words the user is going to say during the application (a Vocabulary).

b. Simple rules about the orders and patterns in which these words may be spoken (a Grammar).

c. A table of computer codes for each word (the Translation Table). Grammar definition sections specify what statements the Voice I/O System can recognize, but not what the system should send to the host computer in response to each recognition. By adding a translation table to the grammar-definition file, the designer can specify the content of a message the Voice I/O System will send to the host computer when a valid statement has been recognized. For example if the word MISSILE which is included into an utterance is recognized, and we want only the letter M to appear on the screen we type the following format of a translation table, for an application containing a single grammar:

```
!grammar_name=

    #recognition

    #grammar

    #translations

    <initiator

    |separator

    >terminator

    MISSILE          M|015.
```

d. Voice Response information which are words and sentences to be spoken by the Recognizers internal speech synthesizer (the Voice Response Facility) when certain statements are recognized.

e. A list of values for certain internal parameters which the user can set in the Recognizer File (called a Setup Block). When the Recognizer File containing a Setup Block is accessed by the Recognizer, all Setup Parameters are set; those listed in the Setup Block are set to the values listed, and all those not listed in the Setup Block are set to their default values. For example we want to set the voice_settings_speed parameter from 5 which is the default value to 6 (allowed values 0 to 9)and the voice_settings_pitch from 5 to 7 (allowed values 0 to 9). The Setup Block will look like this:

```
#set up

voice_settings_speed          = 6

voice_settings_pitch          = 7
```

2. Voice File

The Voice File contains a library of sound patterns for all the words in the Recognizer File; both as they sound when they are spoken individually ("discretely") and as they sound when they are spoken together ("continuously") in the

orders and patterns set forth in the Grammar in the Recognizer File.

Only when the Recognizer has both a Recognizer File and a Voice File can it begin recognition.

Once those files are created, the user must transfer the Recognizer File into one of the sections, or images, in the Recognizer's memory. Once there, it can be stored on a voice cartridge.

Each user must train the Recognizer using the sound of his/her voice. The sound patterns ("Voice File") created through Training can then be stored on the Voice Cartridge along with the Recognizer File.

After the files are stored on a voice cartridge, using the system becomes a one-step process. Whenever the Recognizer is to be used, the user need only switch it on and insert the voice cartridge into the cartridge bay. The voice cartridge contains the Recognizer File and the operator's Voice File.

The Recognizer then automatically begins listening and translating what it hears (VERBEX Installation Manual, 1990). [Ref. 5]

C. DESCRIPTION OF THE PROBLEM

It is a reality today that Naval Operations are done with the use of specially trained personnel due to the complexity of the existing systems such as Tactical Tables for Detection,

Tracking, Recognition, and Establishing Targeting Priorities for the available onboard weapons systems.

These persons have to work under difficult conditions in CIC's (Combat Information Centers) which are usually confined, dimly lit rooms. It is necessary to emphasize that under difficult situations such as rough seas, the abilities of any human are limited to performing only essential functions in order to overcome the difficulties that he/she is facing. This includes every undesirable movement into the operational area or even any repetition of an order.

Under these situations the CO's , XO's, or ever OPS Officers/CIC Officers who are responsible for making the evaluation of a combination of different information passed to them by different sources, may mentally lose a number of their operational personnel.

The situation is further complicated given the fact that after the evaluation process they have to give an order which is required to be executed immediately, and their personnel does not respond fast enough.

Even with the most trained personnel the delay times cannot be eliminated. The need of the undesirable in any case, but impossible to avoid, "middle man" generates a lot of misinterpretation problems which in real time situations may be crucial or even fatal.

Sometimes, because the CO, XO, or the OPS/CIC Officer, away from their positions, are unable to enter data, or their

hands are busy and cannot be used at the specific needed time; this delay may give the enemy the tactical advantage.

It is known that the final decision is always left to the CO of every unit; sometimes this same person has to be on the bridge instead of being in the CIC and, in an emergency case he cannot order immediate action without the need to use one or more "middle men".

It is obvious that the Navies in the whole world are facing the problem of trying to obtain more "real-time" systems, but instead are operating them with the use of the "middle man" as the interface between the decision maker and the computing system.

Using Speech Recognition as that interface will make the computer a true extension of man, solving many problems created by the need of using our hands and eyes in low light conditions with distance limitations.

With the use of this new technology such as the VERBEX Series 5000 Version 3.00, we would not only give operators the advantage of being able to transmit information and operational decisions safely and timely but also will reduce the number of personnel needed to run the available equipment and thus reduce training hours in favor of other activities.

The different features offered by these devices offer the security level needed to limit the number of the operating personnel with the use of the Speech Recognition System to just the essential ones.

III. OBJECTIVE

The objective of this thesis is to investigate the possible use of a Continuous Voice Recognition System (in this case a VERBEX Series 5000 Version 3.00) for inputting operational information to Tactical Tables in Naval Operations.

A secondary objective is measuring the accuracy of the system, how much it may influence our reaction time and, how credible this system is against the procedures currently being used for Tactical Table procedures in Naval Operations.

IV. DESCRIPTION OF THE SIMULATION

A C.I.C. room includes in its systems, two Tactical Tables which do not communicate with each other. The purpose of Tactical Table number one is for assigning guns and torpedoes only to the targets we want to attack, and also has as supplementary features to detect, track, characterize the different targets, and to plot their rackets. Tactical Table number two has the purpose of detecting, tracking, and assigning only missiles to the designated targets.

The two Tactical Tables are distant from each other as shown in Figure 1; so to have the picture of both screens (P.P.I.'S) the CO, XO, or the OPS Officer has to move, in order to insure that exactly the same information has been entered in both systems.

To enter any information in either of the two Tactical Tables, the operator has to use his left hand for the keyboard, and the his right hand for the rolling ball which moves the targeting cross giving at any time, the bearing and range of its position on the screen relatively from the ship.

Operations in the C.I.C. are always done in the dark and the use of eyes as well of hands becomes more difficult, especially when we want to observe specific buttons having special features at the specific time we need them.

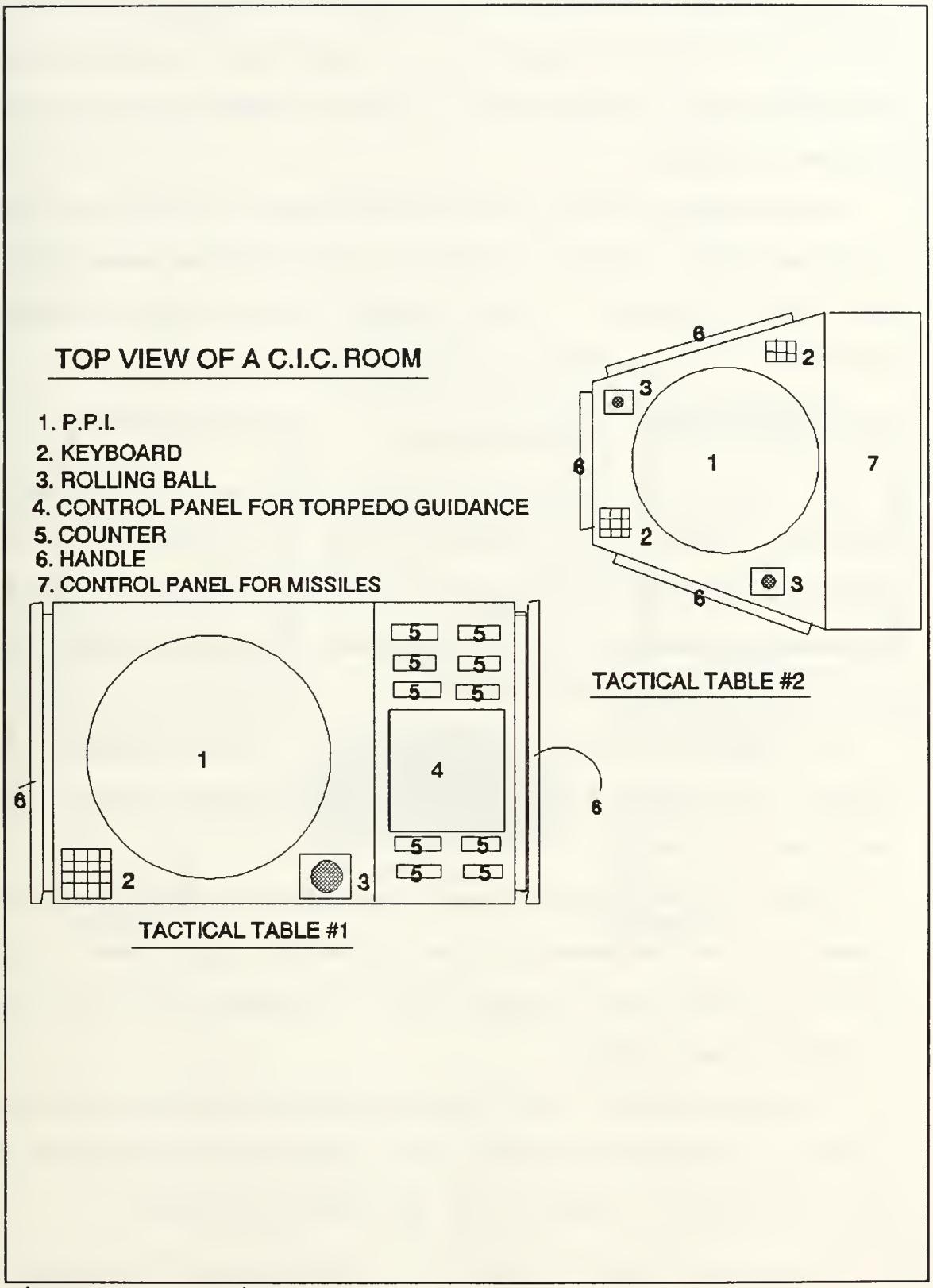

Figure 1: Top View of the C.I.C.

The location of the C.I.C. varies from type to type of ships and the most of the time is located far from the bridge, thus hampering the movements of the CO when he is needed in another location.

Here comes the role of the "middle man" or the role of the intercom system, which provides so many misinterpretations in the voice commands of the decision maker when delegating commands to the operational center.

A. PROCEDURE FOR THIS EXPERIMENT

To simulate the two Tactical Tables, two personal computers (PC's) were connected with a T-Connector to the VERBEX Series 5000 Version 3.00 as shown in Figure 2. The two PC's were four meters apart from each other like in the real C I C environment.

PC-1 had the ability to communicate with the VERBEX Voice System and perform all the features this system offers; PC-2 was used as a monitor.

The voice application as a text file on disk, called a grammar file represented the different functions of the Tactical Tables. The grammar file is created with the use of a standard text editor.

A unique grammar file is necessary for each application of a Speech Recognition System, the contents of which are the data necessary to operate the application program.

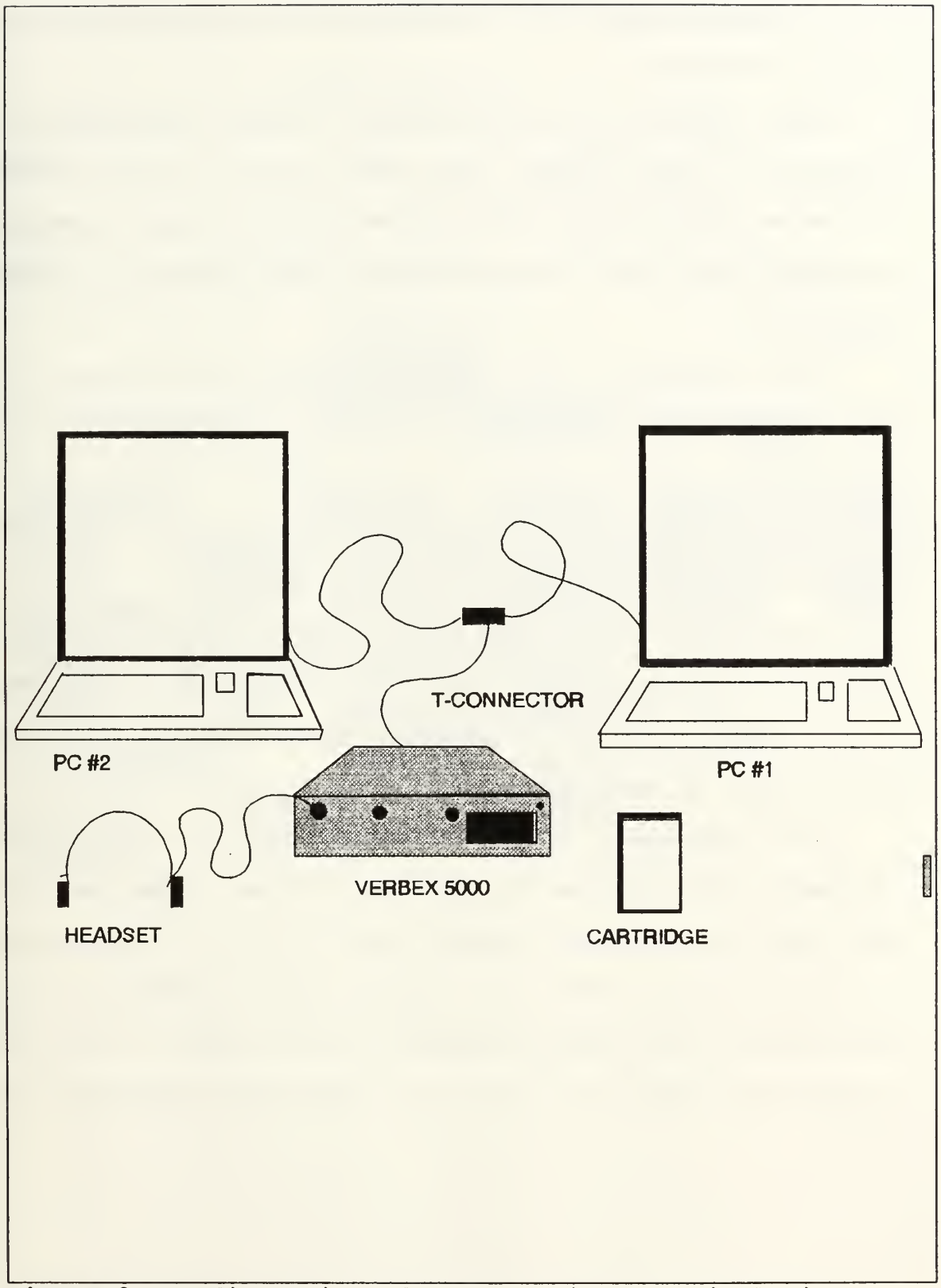

Figure 2: Configuration of the Experimental Simulation.

The grammar file anatomy developed for this application is shown in Figure 3.

Each function of the Tactical Tables represented an utterance in the grammar, and with the use of the CONVERT software, this text file was converted into a binary, machine-readable recognizer file suitable for downloading to a VERBEX Recognizer.

This recognizer file was transferred with the use of a file transferring tool named TRANSFER, to the VERBEX Voice cartridge.

From this moment on a user's voice patterns could be added to the cartridge through the TRAINING process.

Trying to represent the functions of a Tactical Table, as realistically as possible, the author simulated the movements of the rolling ball with spaces in the scenario used for the typing part of the experiment.

The experiment differs from reality in the existing darkness level, because it was done with day light instead of the existing darkness in the ship's C.I.C. Therefore any results obtained from the typing input condition of this experiment would only be expected to be worse in the darker conditions found in a C.I.C. in real world operations of a ship.

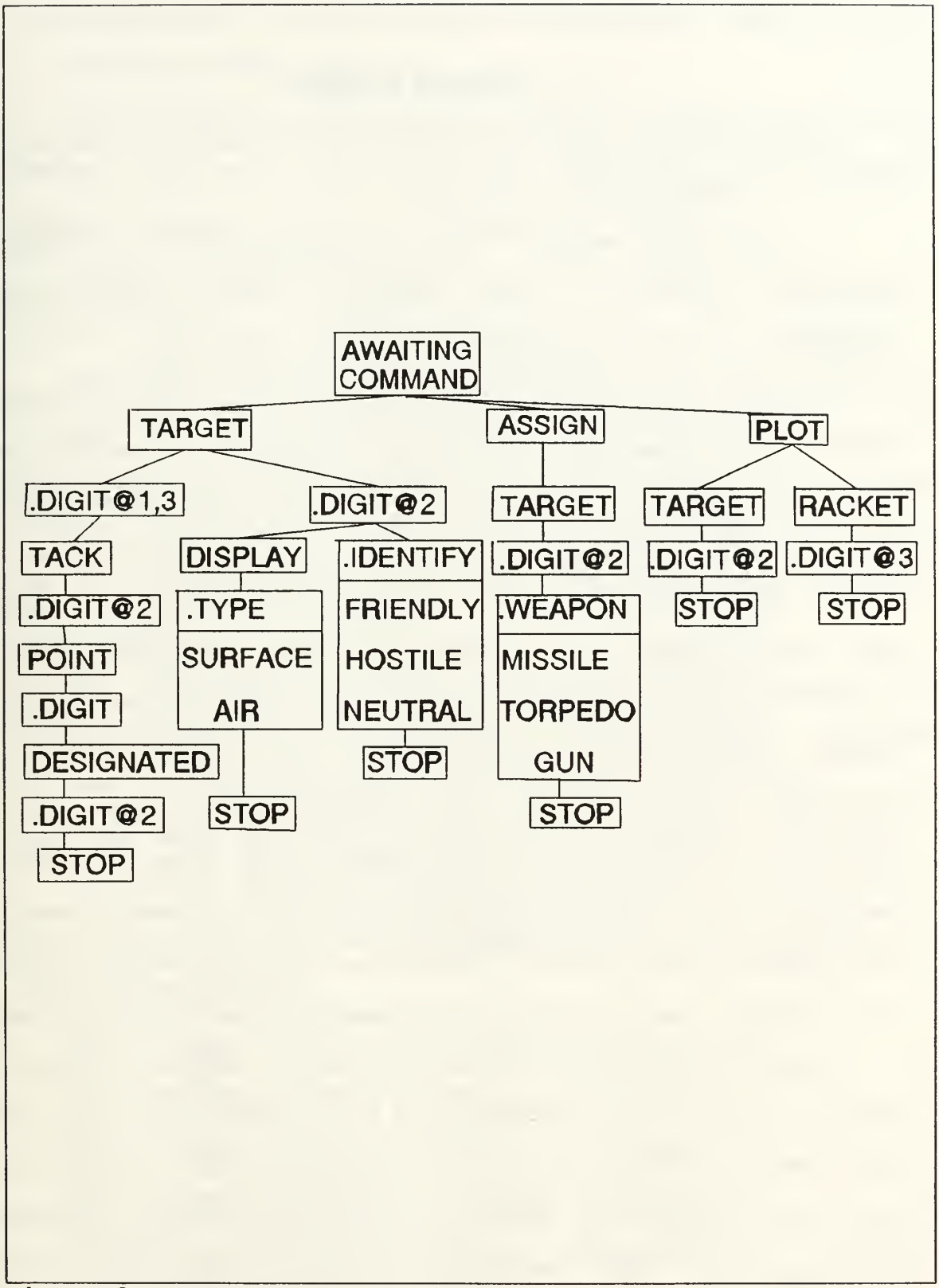

Figure 3: Flow Chart.

V. RESEARCH SCENARIO

This research design was created to simulate reality as close as possible for each case.

As mentioned before, for this experiment the different subjects followed a fixed scenario which comprised two different parts; a written part, and an oral one.

The purpose for having two parts was to illustrate the difference between the time needed to perform a specific task using the VERBEX Series 5000 Version 3.00 as an input mode to the Tactical Tables, and the original manual input mode.

In the scenario, we are operating in the central Aegean Sea, an Archipelagos which make up 2,463 of Greece's total of 3,100 islands. The Aegean Sea as an entity, together with the Greek mainland and the mosaic of the islands which it includes, is an area of vital strategic importance and is absolutely necessary for the defense of Greece. This is because it provides the strategic and tactical depth required for maneuvers and support of military operations and ensures the strategic warning against a preemptive massive air attack.

This area also constitutes successive defense barriers in depth since it is an extension of the Dardanelles Straits and provides, to whoever holds it, the capability to control the sea exits through the Aegean to the Mediterranean. Moreover, in conjunction with the island of Crete, it provides for the

full control of the southeastern Mediterranean region (THREAT in the Agean,1987).[Ref. 6]

In the scenario, the subjects are on board a Hellenic Navy FPBGMT (Fast Patrol Boat with Guns, Missiles, and Torpedoes) and the tactical situation is shown in Figure 4.

Every target which appears on Figure 4 is used the Speech Recognition and in the manual typing part of the scenario.

The experimental design for the manual part was formulated to simulate, as real as possible, the movements made by the operator of the Tactical Table when working manually; and the provided spaces were to simulate the delay times due to the use of the rolling ball with the right hand, which controls the position of the cursor.

The experimental design for the Speech Recognition condition made use of standard Naval terminology to create the utterances in the grammar file. This permitted subjects to use familiar terms and perform the task with minimal trainning.

The voice orders for the Speech Recognition part of the experiment with the use of VERBEX Series 5000 Version 3.00 were the following:

Say: Target 030 Tack 20 Point 0 Designated 01

Say: Target 060 Tack 20 Point 0 Designated 02

Say: Target 090 Tack 25 Point 0 Designated 03

Say: Target 120 Tack 15 Point 0 Designated 04

Say: Target 150 Tack 20 Point 0 Designated 05

Say: Target 180 Tack 10 Point 0 Designated 06

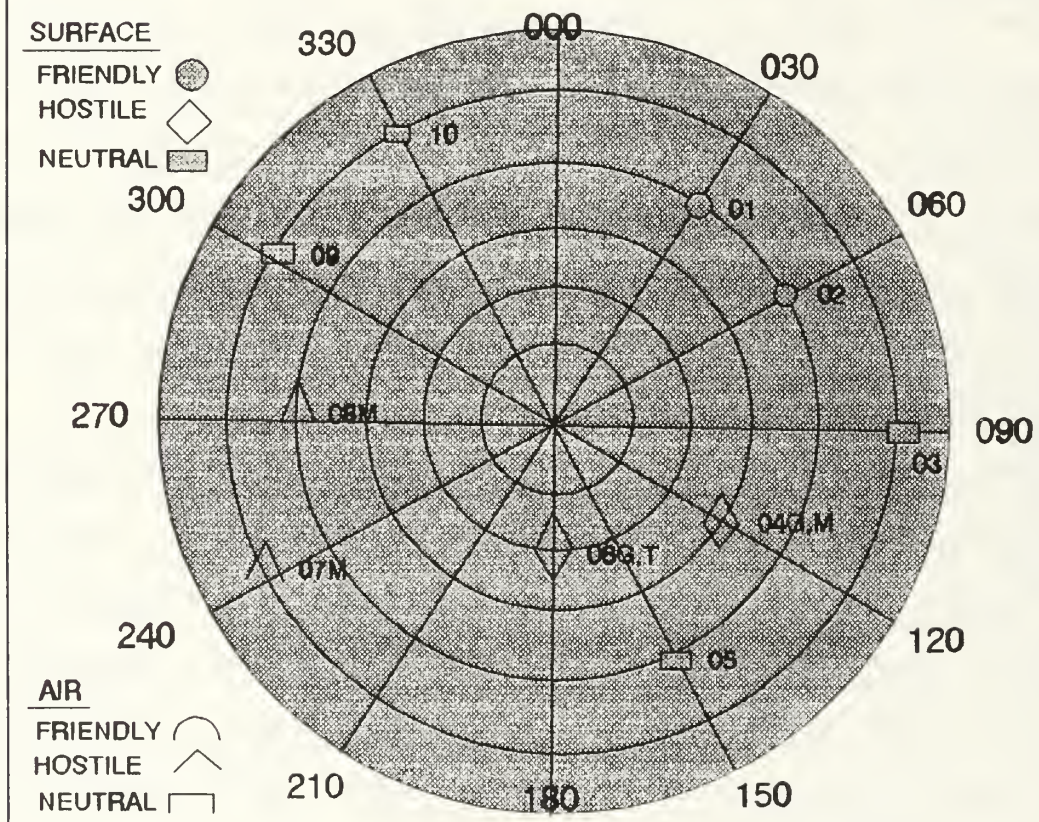

Figure 4: P.P.I. Display of the scenario used.

```
Say: Target 240 Tack 25 Point 0 Designated 07
Say: Target 270 Tack 20 Point 0 Designated 08
Say: Target 300 Tack 25 Point 0 Designated 09
Say: Target 330 Tack 25 Point 0 Designated 10
Say: Target 01 Display Surface
Say: Target 02 Display Surface
Say: Target 03 Display Surface
Say: Target 04 Display Surface
Say: Target 05 Display Surface
Say: Target 06 Display Surface
Say: Target 07 Display Air
Say: Target 08 Display Air
Say: Target 09 Display Surface
Say: Target 10 Display Surface
Say: Target 01 Friendly
Say: Target 02 Friendly
Say: Target 03 Neutral
Say: Target 04 Hostile
Say: Target 05 Neutral
Say: Target 06 Hostile
Say: Target 07 Hostile
Say: Target 08 Hostile
Say: Target 09 Neutral
Say: Target 10 Neutral
Say: Plot Target 04
Say: Plot Target 06
```

Say: Plot Target 07

Say: Plot Target 08

Say: Plot Racket 060

Say: Plot Racket 120

Say: Plot Racket 240

Say: Plot Racket 270

Say: Assign Target 04 Gun

Say: Assign Target 06 Torpedo

Say: Assign Target 07 Missile

Say: Assign Target 04 Missile

The manual inputs for the experiment were the following:

Type: 030-20.0 01

Type: 060-20.0 02

Type: 090-25.0 03

Type: 120-15.0 04

Type: 150-20.0 05

Type: 180-10.0 06

Type: 240-25.0 07

Type: 270-20.0 08

Type: 300-25.0 09

Type: 330-25.0 10

Type: 01 Surface

Type: 02 Surface

Type: 03 Surface

Type: 04 Surface

Type: 05 Surface

```
Type: 06 Surface
Type: 07 Air
Type: 08 Air
Type: 09 Surface
Type: 10 Surface
Type: 01 Friendly
Type: 02 Friendly
Type: 03 Neutral
Type: 04 Hostile
Type: 05 Neutral
Type: 06 Hostile
Type: 07 Hostile
Type: 08 Hostile
Type: 09 Neutral
Type: 10 Neutral
Type: 04 P
Type: 06 P
Type: 07 P
Type: 08 P
Type: 060 R
Type: 120 R
Type: 240 R
Type: 270 R
Type: 04 G
Type: 06 T
Type: 07 M
```

Type: 08 M

Type: 04 M

The words Say, and Type of the above orders were not to be said or typed; they were written in order to help the individuals to understand the beginning of a new utterance in each case respectively.

In other words, the subject when he/she said: "Target 030 Tack 20 Point 0 Designated 01" during the Speech Recognition part as well as he/she typed: "030-20.0 01" in the manual typing part, he/she will get the same result on the Tactical Table which is "01" on the upper right hand quarter of the cursor's position on the P.P.I.

VI. EXPERIMENTAL DESIGN

The experiment was run from 12.00 to 14.00 so the individuals used were not rested after their daily schedule of classes.

Subjects individually met with the experimenter initially and were told about the basic ideas of how the voice recognition equipment worked, what they were expected to say, and how the training on the equipment would be performed. At the same time, they were informed about the simulation and the goal of the experiment in order to have a complete understanding of what they were to do.

After the experiment, they were given a subjective questionnaire regarding their opinions about using voice input versus manual typing input for the use in Tactical Tables.

Each subject trained the system using the Emulate software utility (which makes the Host Computer connected to the Host/Computer port of the VERBEX Recognizer function as though it were an ASCII terminal connected to the User/Terminal port). As part of this process, they accessed the Recognizer's Setup Mode menus and Trained the system with their own voice patterns.

The Recognizer first gave them the words and then the utterances/phrases that were used in the experiment. In this way each person "taught" the Voice Recognition System the

sound of his/her voice speaking words and sample phrases in the Recognizer File.

These sound patterns the Recognizer learns from each subject were then stored along with the Recognizer File on a Voice Cartridge (VERBEX Software Utilities Manual, 1990). [Ref. 7]

Once each subject had stored his voice patterns on the cartridge, he/she then performed the oral part of the experiment by going through the voice order list mentioned in chapter V.

Every time the subject spoke into the microphone, the spoken utterance was displayed on both screens - representing the two Tactical Tables - exactly as they appeared in the grammar definition lines. In other words, the utterances appeared different on the screens, but showed up exactly as it would in the real case on the P.P.I.'s of the Tactical Tables.

The start of reading the list coincided with the start of the chronometer which was stopped when the subject ended the list.

The subject had to enter manually one by one every line contained in the typing list - as mentioned in chapter V. - first in PC-1 and then in PC-2 which was four meters distant from the first one. The experimenter started the chronometer which lasted as long as the individual was running the typing test. Finally, every individual answered the questionnaire.

The scenario was performed one time by each subject using the voice input, and one time using the manual typing input.

The concept was to give the minimum opportunities to the individuals to train the system, in order to get more objective results from the point of view that we can minimize our training time in favor of other activities without risking the level of accuracy of the system.

If under these circumstances the results were positive, then it is obvious that the system is not only faster, but more reliable too.

The total number of individuals used for the experiment were ten; nine men and one woman; and all of them were officers of the U.S.N. and U.S.C.G.

VII. DEPENDENT VARIABLES

During all trials, the following were measured:

1) Time to complete the scenario using the VERBEX Series 50000 Version 3.00 Voice Recognizer as an input mode.

2) Time to complete the scenario with manual input mode.

3) Number of oral input command errors.

4) Number of manual input command errors.

5) Time to complete every utterance included in the grammar file using the oral mode.

6) Time to complete every utterance included in the grammar file using the manual mode.

The author was interested in the number of times the computers were instructed to do something wrong. Therefore, on typing "Target" for example, if the command was typed in wrong, it was counted as one error, whether there was one or several actual keystrokes typed wrong. Similarly for voice input, if a subject spoke the wrong scenario command, the Voice Recognizer may have recognized the voice input correctly, but it would be a wrong command to the Tactical Tables it was counted as an error (Poock, 1980). [Ref. 8] Another error was considered if the subject erroneously pronounced an utterance and the Recognizer either didn't reply, or answered with a wrong command. In other words, the author was interested in a detailed analysis of how

many times one voice utterance might get confused with another, because in a real situation the operator might not have the luxury to afford any mistake.

All the data were selected before the individuals answered the questionnaire. These questions can be found in Appendix I.

VIII. RESULTS

A. RESULTS FOR SCENARIO TIMES

Table I below shows the time taken to perform the set of actions in the scenario for every individual when using the Voice Recognizer and the equivalent times when performing the experiment with Manual input.

Table I. Times for Oral and Manual Input Method.

Individual #	Oral Input Mode			Manual Input Mode		
	min	sec	msec	min	sec	msec
1st	02	41	990	08	31	660
2nd	02	43	020	09	43	020
3rd	03	24	380	02	02	020
4th	02	53	920	10	31	850
5th	02	30	660	10	45	110
6th	02	30	990	11	02	660
7th	03	43	980	10	50	730
8th	03	10	720	02	50	100
9th	02	57	700	09	18	770
10th	02	43	520	11	02	860
Mean Value	02	57	182	09	53	018

As can be seen in Table I, voice input was consistently faster than manual typing input by 3.33 times.

The mean value for the time needed to perform the scenario with voice input is 2.95 min , versus of the 9.88 min needed to perform the same scenario with manual input.

This is a statistically significant difference in favor of voice input and this becomes even more important when we take into account that the subjects had only used voice input for no more than an hour and a half.

There was an improvement in time - seven (7) out of ten (10) times - when performing a second pass immediately after the first one, with the rest of the cases - three (3) out of ten(10) - deteriorating by less than 20 sec compared to their first pass.

On the other hand there was no difference in typing ability with respect to times.

B. RESULTS FOR ERRORS

Figure 5 illustrates the errors input to the system with Oral and Manual Mode. One can see the voice input oral method consistently produced fewer errors.

C. TIMING RESULTS ON INDIVIDUAL UTTERANCES

Table II below shows the times taken for each individual to perform every single utterance included in the grammar file with Oral Mode and Manual Mode respectively.

This is done because in reality the operator does not always have to enter so much operational information to the

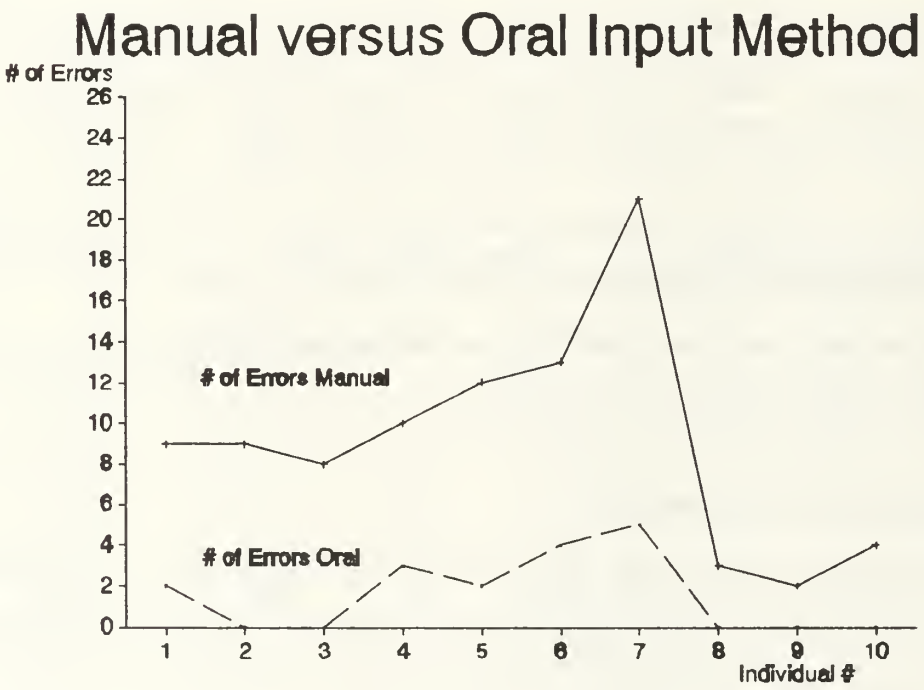

Figure 5: Errors Input to the system with Oral and Manual Mode.

Tactical Tables and in this way we have a more global idea of the timing performance of the compared modes.

Table II. Timing results on individual utterances.

Subject #	Utterance #	Oral Mode		Manual Mode	
		sec	msec	sec	msec
1st	1st	05	690	10	070
	2nd	03	370	08	590
	3rd	02	680	08	790
	4th	02	300	04	610
	5th	02	950	05	790
	6th	02	890	04	820
2nd	1st	07	010	12	870
	2nd	03	600	10	210
	3rd	02	960	10	950
	4th	02	690	05	780
	5th	03	150	06	890
	6th	03	120	05	930

3rd	1st	06	740	15	910
	2nd	03	820	08	050
	3rd	03	310	08	640
	4th	03	090	05	340
	5th	03	620	06	690
	6th	03	810	04	810
4th	1st	05	330	18	610
	2nd	03	380	14	140
	3rd	02	620	13	870
	4th	02	690	06	450
	5th	02	990	08	860
	6th	03	160	07	010
5th	1st	07	020	18	330
	2nd	03	700	15	140
	3rd	02	570	12	870
	4th	02	890	06	450
	5th	03	250	09	330
	6th	03	110	07	100

6th	1st	05	100	14	820
	2nd	03	090	12	840
	3rd	02	410	10	800
	4th	02	500	05	720
	5th	02	840	06	660
	6th	03	030	07	230
7th	1st	05	750	13	910
	2nd	03	080	08	630
	3rd	02	440	09	280
	4th	02	450	04	890
	5th	02	860	06	240
	6th	02	800	05	160
8th	1st	05	220	13	220
	2nd	03	020	07	880
	3rd	02	780	08	700
	4th	02	510	06	710
	5th	02	730	06	270
	6th	02	830	05	910
9th	1st	06	010	13	230
	2nd	03	460	09	630
	3rd	02	700	09	920
	4th	02	660	05	230
	5th	03	030	08	180
	6th	03	220	05	100

10th	1st	06	560	14	720
	2nd	03	410	11	160
	3rd	02	590	12	050
	4th	02	630	06	630
	5th	03	070	09	030
	6th	03	360	07	45

To clarify the meaning of the "utterance #" which is comprise into the Table II, the 1st utterance for the Oral Mode is of the type TARGET .DIGIT @1,3 TACK .DIGIT @2 POINT .DIGIT DESIGNATED .DIGIT @2, and .DIGIT @3 - .DIGIT @2 . .DIGIT .DIGIT@2 for the Manual Mode.

Respectively the 2nd utterance is of the type TARGET .DIGIT @2 DISPLAY .TYPE for the Oral Mode and .DIGIT @2 .TYPE for the Manul Mode. The 3rd utterance is of the type TARGET .DIGIT @2 .IDENTIFY for the Oral Mode and .DIGIT @2 .IDENTIFY for the Manual Mode.

The 4th utterance is of the type PLOT TARGET .DIGIT @2 for the Oral Mode and .DIGIT @2 P for the Manual Mode. The 5th utterance is of the type PLOT RACKET .DIGIT @3 for the Oral Mode and .DIGIT @3 R for the Manual Mode.

Finally, the 6th utterance is of the type ASSIGN TARGET .DIGIT @2 .WEAPON for the Oral Mode and .DIGIT @2 G/T/M for the Manual Mode.

D. SUBJECTIVE QUESTIONNAIRE RESULTS

As mentioned above, after completing their performance with the Oral and Written part of the experimental scenario, subjects filled out a questionnaire with the following results:

(1). When asked if they had any familiarity with Tactical Data Systems, fifty percent answered NO.

(2). When asked if they had any familiarity with Voice Recognition Systems, eighty percent of them answered YES. Their familiarity was a result of the OS-3404 course they have taken in the N.P.S. during their coursework to fulfill the requirements necessary for a Master of Science in Telecommunications System Management.

(3). When asked if they believed that a Continuous Voice Recognition System would be useful as a method of Voice Input versus Manual keying, all of them answered positively, saying also that the operator is faster, has his hands free,and he can concentrate better on his/her job.

(4). When asked if they believed that a Voice Recognition System as an Input tool for a Tactical Data System would help increase the combat reaction time available to the Tactical Commander, all of the responses were positive.

IX. OTHER OBSERVATIONS

A. Several subjects mentioned that with the Voice Input they felt they had better control of the situation. They had their hands free and did not need to concentrate on typing the correct command rather than observing what they were doing.

B. After the fifth individual and during the execution of the experiment, the author noticed an increasing number of misrecognitions (not errors) with the Voice Input Mode. After cleaning the contacts (pins) of the Cartridge the number of misrecognitions had dropped to zero for the following two experiments.

C. Instead of the standalone (independently housed) voice processor unit of the VERBEX 5000, it would be desirable to build the processor into the Tactical Table. This is technologically feasible, since voice processor add-in boards are currently available for the IBM PC and other microcomputers. All that needs to be done is to interface the voice processor circuitry to the Tactical Table. This would save space as well as eliminating the need to check the Cartridge to ensure it is in the right position in case of rough seas. It is important to note that military specifications are beyond of what is offered in this version of the VERBEX 5000. Another point equally important to mention is that it would be more realistic if the Cartridge could

store the voice patterns of all operators so that it did not need to be re-initialized each time a different user came on watch during real operations.

D. Speech Recognition Systems, while they generally do not affect the internal workings of the computer, present their greatest potential advantage in increasing the efficiency of the total human/machine system. Computer input by voice allows those who are not familiar or comfortable with the computer to comprehend and follow what the operator may see unfolding at his work station, without requiring lengthy and distracting explanations from the operator (French, 1983). [Ref. 9].

During the experiments it was noticed that the subjects exhibited signs of stress, even though the author didn't impose on them any kind of stress by limiting the time to perform the experiment. Another factor which appeared to induce stress upon the subjects was the presence of the author during the experiment.

Upon completion of the execution of the Voice Input Mode most of the subjects complained that they were not very familiar with typing. They also inquired as to how well the other subjects were performing the Manual Input Mode exercises.

When using the Voice Recognition Mode, subjects with symptoms of stress appeared to talk in longer bursts, with shorter pauses separating the bursts. Because psychological

stress has a definite effect on the voice, it is reasonable to expect it to have a negative effect on the success rates of users of voice input equipment (French, 1983).[Ref. 9]

E. Experimenting further with one subject, the author noticed that when working with only one Tactical Table - one PC in our case - and using the longest of the Speech Recognition utterances, the time needed to perform the task was 5 sec 950 msec. The time for the same subject to perform the same utterance in the manual typing mode was measured to be slightly faster, 5 seconds 020 milli-seconds when he knew from the beginning what he was to type. On the other hand when he didn't know the exact content of the utterance, the time he took to perform the same task manually was 7 seconds 740 milli-seconds.

X. CONCLUSIONS

In summary, it can be said that the use of a Continuous Voice Recognition System for inputting operational information to Tactical Tables in Naval Operations is feasible. For the experiment, results show that it is faster - by 3.3 times - than the manual typing input and at the same time more accurate by 5.69 times. In this way operators can decrease their reaction time in a real situation with minimal possible errors.

Operators can decrease their dependence on factors such as darkness and available space to move to enter data, as well minimize their dependence on the cooperation of others. During peace time supervisors can reduce the operators' training hours in favor of other activities.

On the other hand, to increase the credibility of such a system it will be necessary to create a recognizer file which is able to provide a feedback response back to the user before the execution of crucial orders such as to fire a missile on a previously defined target before the launch. This of course could further decrease the reaction time of the system.

APPENDIX A

EXPERIMENTAL DATA SHEET (QUESTIONNAIRE)

1. M_____ F_____

2. AGE_____

3. SERVICE_____

4. FAMILIARITY WITH TACTICAL DATA SYSTEMS Y_____ N_____

   IF YES, WHICH ONE(S) ?

5. FAMILIARITY WITH VOICE RECOGNITION SYSTEMS Y____ N____

   IF YES, WHICH ONE(S) ?

6. DO YOU BELIEVE THAT A CONTINUOUS VOICE RECOGNITION SYSTEM WOULD BE USEFUL AS A METHOD OF VOICE INPUT VERSUS MANUAL KEYING ? WHY OR WHY NOT ?

7. DO YOU BELIEVE THAT VOICE RECOGNITION AS AN INPUT TOOL FOR A TACTICAL DATA SYSTEM WOULD HELP TO INCREASE THE COMBAT REACTION TIME AVAILABLE TO THE TACTICAL COMMANDER ? WHY OR WHY NOT ?

APPENDIX B

GRAMMAR FILE USED FOR THE EXPERIMENT

```
!TACTABLE_GRAM=
#REC
#G
 TARGET .DIGIT @1,3 TACK .DIGIT @2 POINT .DIGIT
& DESIGNATED .DIGIT @2
 TARGET .DIGIT @2 DISPLAY .TYPE
 TARGET .DIGIT @2 .IDENTIFY
 PLOT RACKET .DIGIT @3
 PLOT TARGET .DIGIT @2
 ASSIGN TARGET .DIGIT @2 .WEAPON
.DIGIT=
    0
    1
    2
    3
    4
    5
    6
    7
    8
    9
```

```
.TYPE=
    SURFACE
    AIR
.IDENTIFY=
    FRIENDLY
    HOSTILE
    NEUTRAL
.WEAPON=
    MISSILE
    TORPEDO
    GUN
#TR
TARGET                          |040
TACK                            -
POINT                           .
DESIGNATED                      DESIGNATED
SURFACE                         SURFACE |015
AIR                             AIR |015
DISPLAY                         |040
IDENTIFY                        |040
FRIENDLY                        FRIENDLY |015
HOSTILE                         HOSTILE |015
NEUTRAL                         NEUTRAL |015
PLOT                            |040
RACKET                          RACKET
MISSILE                         M|015
```

TORPEDO	T|015
GUN	G|015

LIST OF REFERENCES

1.Strategic Computing Program, Integration, Transition and Performance Evaluation of Speech Technology, Draft, Chapter 2, December 1985.

2.LeFever, Michael A., Speech Recognition in a Command and Control Workstation Environment, Naval Postgraduate School Master's Thesis, Monterey, CA, March 1987.

3.Poock, Gary K., Speech Recognition Research, Applications and International Efforts, Naval Postgraduate School, Monterey, CA, Proceedings of the Human Factors Society Annual Conference, Dayton, Ohio, 1986.

4.VERBEX Conversational Voice Input/Output System GRAMMAR DEVELOPMENT MANUAL, Revision 1.03, NJ, March 1990.

5.VERBEX SERIES 5000 Conversational Voice Input/Output System, INSTALLATION MANUAL, Revision 3.00, NJ, March 1990.

6.The Journalists' Union of the ATHENS Daily Newspapers, THREAT in the Aegean, GREECE, 1987.

7.VERBEX Conversational Voice Input/Output System, SOFTWARE UTILITIES MANUAL, Revision 3.0, NJ, February 1990.

8.Poock, Gary K., Experiments with Voice Input for Command and Control: using voice input to operate a Distributed Computer Network, Naval Postgraduate School, Monterey, CA, April 1980.

9.French, Brian A., Some effects of stress on users of a Voice Recognition System: a preliminary inquiry, Naval Postgraduate School Master's Thesis, Monterey, CA, March 1983.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Technical Information Center Cameron Station Alexandria, VA 22304-6145	2
2.	Library, Code 52 Naval Postgraduate School Monterey, CA 93943-5002	2
3.	Dr. Gary K. Poock, OR/Pk Department of Operations Research Naval Postgraduate School Monterey, CA 93943-5002	2
4.	Dr. Tung X. Bui, AS/Bd Department of Administrative Sciences Naval Postgraduate School Monterey, CA 93943-5002	1
5.	Embassy of Greece Naval Attache 2228 Massachusetts Ave., N.W. Washington, D.C. 20008	5
6.	Constantinos P. Leventis Alevizatou 82 Papagou 15669 Athens, Greece	3